Reliable, robust and realistic: the three R’s of next-generation land-surface modelling

I. C. Prentice¹,², X. Liang³, B. E. Medlyn²,⁴, and Y.-P. Wang⁵

¹AXA Chair of Biosphere and Climate Impacts, Grand Challenges in Ecosystems and the Environment and Grantham Institute – Climate Change and the Environment, Department of Life Sciences, Silwood Park Campus, Ascot, Imperial College London, UK
²Department of Biological Sciences, Macquarie University, North Ryde, New South Wales 2109, Australia
³Department of Civil and Environmental Engineering, University of Pittsburgh, Pennsylvania, USA
⁴Hawkesbury Institute for the Environment, University of Western Sydney, Locked Bag 1797, Penrith, New South Wales, Australia
⁵CSIRO Ocean and Atmosphere Flagship, Private Bag 1, Aspendale, Victoria, Australia

Correspondence to: I. C. Prentice (c.prentice@imperial.ac.uk)

Received: 28 July 2014 – Published in Atmos. Chem. Phys. Discuss.: 26 September 2014
Revised: 9 April 2015 – Accepted: 13 April 2015 – Published: 29 May 2015

Abstract. Land-surface models (LSMs) are increasingly called upon to represent not only the exchanges of energy, water and momentum across the land–atmosphere interface (their original purpose in climate models), but also how ecosystems and water resources respond to climate, atmospheric environment, land-use and land-use change, and how these responses in turn influence land–atmosphere fluxes of carbon dioxide (CO₂), trace gases and other species that affect the composition and chemistry of the atmosphere. However, the LSMs embedded in state-of-the-art climate models differ in how they represent fundamental aspects of the hydrological and carbon cycles, resulting in large inter-model differences and sometimes faulty predictions. These “third-generation” LSMs respect the close coupling of the carbon and water cycles through plants, but otherwise tend to be under-constrained, and have not taken full advantage of robust hydrological parameterizations that were independently developed in offline models. Benchmarking, combining multiple sources of atmospheric, biospheric and hydrological data, should be a required component of LSM development, but this field has been relatively poorly supported and intermittently pursued. Moreover, benchmarking alone is not sufficient to ensure that models improve. Increasing complexity may increase realism but decrease reliability and robustness, by increasing the number of poorly known model parameters. In contrast, simplifying the representation of complex processes by stochastic parameterization (the representation of unresolved processes by statistical distributions of values) has been shown to improve model reliability and realism in both atmospheric and land-surface modelling contexts. We provide examples for important processes in hydrology (the generation of runoff and flow routing in heterogeneous catchments) and biology (carbon uptake by species-diverse ecosystems). We propose that the way forward for next-generation complex LSMs will include: (a) representations of biological and hydrological processes based on the implementation of multiple internal constraints; (b) systematic application of benchmarking and data assimilation techniques to optimize parameter values and thereby test the structural adequacy of models; and (c) stochastic parameterization of unresolved variability, applied in both the hydrological and the biological domains.

1 Introduction

The land surface, together with the soil column underneath it, plays a key role in controlling not only the partitioning of available energy (into latent, sensible and ground heat fluxes) and water (into evapotranspiration, surface runoff, interflow, baseflow and soil moisture), but also the land–atmosphere exchange of carbon dioxide (CO₂) and the close coupling

Published by Copernicus Publications on behalf of the European Geosciences Union.
between photosynthesis and the cycling of energy and water vapour. Adequate representations of biological, physical and hydrological processes in a land-surface model (LSM) are therefore a prerequisite for improving the accuracy of both numerical weather forecasts and climate predictions. LSMS also provide a valuable tool to assess water resources, and the hydrological impacts of changes in climate and land use, over large river basins and continents, having the advantage of a globally consistent physical basis (Eagleson, 1986; Harrison et al., 1991). Moreover, LSMS are being required to perform new functions. In emerging Earth system models, they are called upon to model land–atmosphere exchanges of biogenic greenhouse gases other than CO2; other reactive trace gases with influences on atmospheric chemistry and composition; emissions of aerosols in biomass burning and dust deflation; and emissions of volatile organic compounds as aerosol precursors. This list could be continued, and is lengthening as knowledge increases about the diversity and complexity of Earth system interactions and feedbacks (Friedlingstein et al., 2013; Scholze et al., 2013; Ciais et al., 2014).

Many LSMS now include representations of the slower processes of vegetation dynamics, coupled to the fast exchanges of water, energy, momentum and CO2 that are at their core (Arora, 2002). Dynamic global vegetation models (DGVMs) have been reviewed elsewhere (e.g. Prentice et al., 2007; Tang and Bartlein, 2008; Prentice and Cowling, 2013). Some offline DGVMs (i.e. models not coupled to a climate model) have been used to address water resources questions (e.g. Rost et al., 2008; Murray et al., 2011, 2012a, b). Thus the boundaries between LSMS, DGVMs and global hydrological models are increasingly blurred. Here we focus on LSMS sensu stricto but our treatment applies equally to the representation of core land-surface processes in DGVMs. We first briefly review the evolution of land-surface modelling, then proceed to consider the present state of the art and how it could be improved upon.

The three R’s of the title are all generally recognized as important characteristics of a numerical model, but models often do not possess all three. Possession of one feature does not by any means guarantee the rest. By reliable, we mean a model that gives approximately correct predictions under most circumstances. By robust, we mean a model whose results do not depend sensitively on the specification of quantities that are poorly known. By realistic, we mean a model that includes sufficient processes, represented in adequate detail, to allow simulation of the system’s response to a change in all of the external variables of interest. We will argue that the dominant paradigm in land-surface modelling focusses too heavily on realism at the expense of the other two R’s.

2 Evolution of land-surface models

Land-surface modelling consists of the development and application of computational models integrating biological, hydrological, and physical processes within the soil–plant–atmosphere continuum. LSMS have two essential characteristics: (1) they consider processes related to the energy, water, and carbon cycles and their interactions, and (2) they operate over relatively large spatial domains with short temporal scales. Depending on their complexity, different LSMS may consider different processes and represent them differently.

Manabe (1969) was the first to include land-surface interactions explicitly in a climate model. Manabe’s so-called bucket model includes vastly simplified hydrology (for example, no surface runoff is generated until the entire soil column reaches saturation), a simple energy balance equation, and no explicit vegetation characteristics. But Manabe’s pioneer work ignited many significant developments in later LSMS.

In common with several earlier reviews including the influential article by Sellers et al. (1997), we consider the subsequent evolution of LSMS as a sequence of “generations”, with Manabe’s bucket model representing the first generation. But whereas Sellers et al. (1997) focused exclusively on LSMS as a component of climate models, our treatment also covers the extensive offline development of LSMS for hydrological applications that took place from the late 1980s onwards.

The pioneers of the second generation of LSMS were Deardorff (1978), Dickinson et al. (1986, 1993) (the BATS model) and Sellers et al. (1986, 1996) (the SiB model). These “generation 2A” LSMS focused on achieving a much more detailed representation of vegetation as the locus of many of the physical exchanges between land and the atmosphere, and a more realistic computation of the surface energy budget (Fig. 1). Later models followed along similar lines, including a variety of innovative components (e.g. Noilhan and Planton, 1989; Xue et al., 1991; Koster and Suarez, 1992; Ducoudré et al., 1993; Verseghy et al., 1993; Viterbo and Beljaars, 1995; Wetzel and Boone, 1995; Desborough and Pitman, 1998).

Parallel developments in offline models (Fig. 2) tackled problems caused by the unresolved (subgrid-scale) variability of precipitation and land-surface characteristics (topography, vegetation and soils). Because of the extreme non-linearity of many key processes, disregarding this variability can lead to substantially incorrect computations of the aggregate surface water and energy budgets (e.g. Chen et al., 1997). Stochastic parameterizations, discussed in more depth later, were introduced as a means to deal with this problem of subgrid-scale variability. Attention was also paid to improving the representation of specific hydrological processes including infiltration, surface and subsurface runoff, and processes associated with snow. Representative LSMS in this “generation 2B” include the VIC (Liang et al., 1994, 1996a, b; Liang and Xie, 2001), TOPLATS (e.g. Famiglietti
Figure 1. Schematic of “generation 2A” LSMs. The energy budget is represented by short-wave radiation ($R_s$), long-wave radiation ($R_l$), latent heat flux (LH), sensible heat flux (SH) and ground heat flux ($G$). The water budget is represented by $P$ (precipitation), $E$ (bare ground evaporation), $E_t$ (transpiration), $E_c$ (evaporation from canopy interception) and surface runoff ($R$). The water budget is coupled with the energy budget, but hydrological processes are represented very simply; for example, subsurface runoff is represented only by vertical drainage ($q$). Precipitation, vegetation type and soil properties are treated as constant within each grid cell.

Figure 2. Schematic of “generation 2B” LSMs. See Fig. 1 for basic symbols. In addition, subgrid variabilities of precipitation, vegetation type, soil properties and topography are represented statistically ($\mu$ represents the variable precipitation-covered area) and hydrological processes are represented more explicitly. Thus surface and subsurface runoff ($Q_b$ and $q$) are distinguished, and diffusion ($D$), lateral flow in the subsurface ($Q_b$), and groundwater table dynamics are also modelled.

Figure 3. Schematic of “third-generation” LSMs, which are similar to generation 2A (Fig. 1) except that now the carbon budget is coupled to the calculation of the water and energy budgets through parameterizations of stomatal behaviour. However, these models do not incorporate the improved treatments of subgrid spatial variability and hydrological processes developed in generation 2B (Fig. 2).
on the relative carbon “costs” of maintaining the water flow pathway required for transpiration and the biochemical capacity for photosynthesis.

Representing land–atmosphere exchanges of water and carbon also required a representation of dynamic changes in green vegetation cover, especially the seasonal cycle. But how to represent vegetation phenology in a model is still a work in progress. Two principal approaches can be distinguished: plant-physiological (e.g. Lu et al., 2001) and rule-based (e.g. Foley et al., 1996; Levis and Bonan, 2004; Kim and Wang, 2005). This remains one of the least well modelled aspects of the land surface (Keenan et al., 2014). One promising avenue of development considers the biologically adaptive nature of phenology (Caldararu et al., 2014), leading to the idea of biologically optimized control of leaf flushing and senescence.

Many LSMs are now coupled to explicit representations of vegetation dynamics, represented by quantitative mixtures of plant functional types (PFTs) that are updated at intervals much longer than the time step of the LSMs. The land-surface component of many climate and Earth system models is therefore now a full DGVM, representing a cascade of processes with intrinsic time scales ranging from minutes to centuries, with asynchronous coupling to link faster and slower processes (Prentice et al., 2007). This development could, optimistically, be regarded as a major achievement in the integration of physical and biological aspects of the land surface (McGill et al., 2006). However, as discussed in the next section, the performance of such models has proved inconsistent. Reliability appears to have been lost in the scramble to develop multifunctional LSMs. Furthermore, the third-generation models and DGVMs have generally not fully capitalized on advances in the representation of subgrid-scale heterogeneity and hydrological processes made in the second generation. The time is ripe for a synthesis of these elements.

3 Model comparisons, evaluations, and the need for benchmarking

The Programme for Intercomparison of Land-surface Parameterization Schemes (PILPS) was founded in the early 1990s (Henderson-Sellers et al., 1993, 1995) as an attempt to make sense of large differences that had been noted in the behaviour of contemporary LSMs, through community involvement in standardized model “experiments”. The specific goal of PILPS was to improve understanding and implementation of first- and second-generation LSMs, as used to represent land-surface physical processes at regional to continental scales.

PILPS was one of six international efforts later subsumed under the umbrella of the Global Land/Atmosphere System Study (GLASS). GLASS aims to improve model representations of land-surface states and fluxes, to better understand interactions of the land surface with the overlying atmosphere, and to maximize the fraction of inherent predictability in land–atmosphere coupled processes (van den Hurk et al., 2011). PILPS has been through five phases: documenting the status of LSMs (Phase 0), performing offline tests of LSMs using synthetic atmospheric forcings (Phase 1a–c), using observed forcings and observations to evaluate the performance of LSMs offline (Phase 2a–e), coupling tests of LSMs within the Atmospheric Model Intercomparison Project (AMIP) (Phase 3), and evaluation of the performance of LSMs when coupled to their host climate models (Phase 4) (Henderson-Sellers et al., 1996). Results of “point” and small-area studies from PILPS 1a–c and 2a, b and d revealed large differences among models, and the fact that many diverged considerably from observations (e.g. Shao and Henderson-Sellers, 1995; Henderson-Sellers et al., 1996; Chen et al., 1997; Schlosser et al., 2000). PILPS 2c and 2e were carried out for large river basins: 2c focusing on the mid-latitude Red–Arkansas River basin in the central USA, 2e on high-latitude Torne–Kalix basin in Sweden. The principal findings (Liang et al., 1998; Lohmann et al., 1998a; Wood et al., 1998; Bowling et al., 2003a, b; Niessen et al., 2003) were as follows. (1) LSMs that applied subgrid-scale runoff parameterizations could simulate large-scale river discharges better than others. (2) The modelled partitioning between surface and subsurface runoff varied even more than the modelled total runoff. In particular, the runoff parameterizations of LSMs under dry conditions were found to need improvement (Lohmann et al., 1998b; Bowling et al., 2003a). (3) The attenuation of solar short-wave radiation by vegetation needs to be considered in order to calculate the ground heat flux properly (Liang et al., 1998). (4) The partitioning of water and energy (i.e. the modelling of runoff and evapotranspiration) differed greatly among LSMs, even on an annual and monthly basis and even when the same forcing data, vegetation and soil information, and model parameters were used. (5) Mean values and spatial patterns of net radiation and surface temperature in warm conditions generally showed the best agreement among the LSMs, and with observations (Liang et al., 1998). (6) Models that conducted calibrations on some of their parameters performed consistently better than those that did not, regardless of the specific calibration method used. (7) Some model parameters in LSMs were found to be particularly critical for the partitioning of water and energy. For example, in the high-latitude study (PILPS 2e), it was shown using a simple “equivalent model” that variations in the partitioning of precipitation and energy at an annual scale could be attributed primarily to parameters related to snow albedo, effective aerodynamic resistance and evaporation efficiency (Bowling et al., 2003b).

For the mid-latitude study (PILPS 2c), Liang and Guo (2003) applied the fractional factorial method to ten LSMs in order to investigate the sensitivities of four quantities (annual evapotranspiration, total runoff, sensible heat flux, and soil moisture), and their combined effects, to five parameters that the models had in common: maximum soil

moisture content (MSMC), effective available water content, the Clapp–Hornberger $B$ parameter, leaf area index, and minimum stomatal resistance. It was shown that MSMC and the Clapp–Hornberger $B$ were usually the most critical. This study also indicated that variations associated with soil properties (due to measurement uncertainties, and/or spatial heterogeneity) played a stronger role in the partitioning of water and energy budgets than those associated with vegetation properties. Sensitivities to different parameters were found to vary across hydroclimates, and generally the effects of different parameterizations were greater under arid than moist conditions (also shown by Lohmann et al., 1998a).

Despite the achievements of PILPS, and subsequent projects with more specific goals including GSWP (Global Soil Wetness Project: Dirmeyer et al., 1999, 2006), GLACE (Global Land Atmosphere Coupling Experiment: Koster et al., 2004, 2010) and LUCID (Land-Use and Climate, IDENTification of robust impacts: Pitman et al., 2009), many of the most general questions originally posed are still unanswered. This situation was articulated in a review of GLASS by van der Hurk et al. (2011). For example, it is still not clear to what extent predictability can be achieved in a LSM; what parameterizations are more appropriate, under what conditions; and what is the best strategy to reduce prediction uncertainties. Moreover, many of the differences among LSMs, and discrepancies between LSMs and observations, have not been resolved and remain incompletely understood.

The coordinated international activities described above focused on the comparison and evaluation of LSMs sensu stricto. The international LAnd Model Benchmarking (iLAMB) project was inaugurated in 2009 with the explicit goal of a unified approach to the comparison and evaluation of land models including both carbon and water cycling aspects, and an unstated one, to rekindle apparently flagging land models including both carbon and water cycling aspects. The project's stated goals are to (quoted from http://www.ilamb.org/, accessed 20 April 2014):

1. “to develop internationally accepted benchmarks for land model performance,
2. promote the use of these benchmarks by the international community for model intercomparison,
3. strengthen linkages between experimental, remote sensing, and climate modelling communities in the design of new model tests and new measurement programs, and
4. support the design and development of a new, open source, benchmarking software system for use by the international community.”

These goals set out exactly what is required in order to make systematic testing against observations into a routine part of model development. However, the most recent iLAMB workshop took place in January 2011, and the stated goals seem to be some way from achievement. Some groups have published “first draft” sets of benchmark protocols and metrics (Randerson et al., 2009; Kelley et al., 2013) principally (not exclusively) focused on the carbon-cycle aspects. The Protocol for the Analysis of Land-Surface models (PALS) software (Abramowitz, 2005) allows rapid comparison of modelled and observed CO$_2$ and latent heat fluxes at the publicly available eddy-covariance flux measurement stations in the FLUXNET archive. The ecosystem Modelling And Scaling infrasTructure (eMAST) project of the Australian Terrestrial Ecosystem Research Network (TERN) (http://www.tern.org.au/) is assembling diverse data sets and developing software to facilitate terrestrial ecosystem data–model comparison and integration, with an initial focus on the Australian continent. This is by no means a comprehensive list of such initiatives. Nevertheless, our impression is that there is still limited momentum in the coordinated development of international benchmark systems, and that this is to the detriment of LSM improvement.

In summary, the development of LSMs in the climate modelling context has been characterized by intermittent and insufficient attention to model evaluation (Prentice, 2013). Probably as a direct consequence, those aspects of climate model predictions of the historical observational record that depend most strongly on the land-surface component are subject to remarkably large differences between models, which affect the quantification of both climate feedbacks (Ciais et al., 2014) and impacts with major consequences for human society (Schellnhuber, 2014). Two such areas of major disagreement among models were highlighted in the IPCC Fourth Assessment Report (Denman et al., 2007), and persisted without resolution into the Fifth:

(a) The hydrological cycle, specifically the degree to which precipitation over the continents depends on soil moisture and evapotranspiration from the land surface. The GLACE-1 experiment (Koster et al., 2002) showed that different GCMs behave very differently in this respect. Although the differences could be partly due to different schemes for generating precipitation in the atmosphere, the evidence points to differences among LSMs as a prime suspect.

(b) The carbon cycle, specifically the degree to which the growth rate of CO$_2$ in future is likely to be reduced due to enhancement of NPP (“CO$_2$ fertilization”: a negative feedback), and also the extent of compensating increase due to the acceleration of soil organic matter decay in a warming climate (a positive feedback). In the Coupled Carbon-Climate Model Intercomparison Project (C$^4$MIP) (Friedlingstein et al., 2006) the participating models agreed that the sign of the feedback from climate change to atmospheric CO$_2$ is positive, i.e. the effect of a warming climate is to release
CO₂ from the land surface. Some new models including C–N cycle coupling have predicted the opposite sign, i.e. a negative feedback (Thornton et al., 2007; Sokolov et al., 2008), although this is not consistent with evidence from past changes in atmospheric CO₂ concentration shown in ice-core records of the past millennium (Friedlingstein et al., 2010). The models reported in the IPCC Fifth Assessment Report (AR5) have produced carbon–climate feedbacks with consistently positive sign, but varying greatly in magnitude (Ciais et al., 2014). Most of the AR5 models underestimate the historical CO₂ uptake by ocean and land (Hoffman et al., 2014). A model comparison against two Free Air Carbon dioxide Enrichment (FACE) experiments (Zaehle et al., 2014) found that the land C cycle component of one model in AR5 that includes a representation of C–N cycle coupling (CLM4) systematically underestimated the observed response of NPP to CO₂ enhancement.

The differences among different models’ predictions of 21st century CO₂ uptake have remained large through successive IPCC Assessments (Fig. 4). Alarming, the spread of modelled present values of gross primary production (GPP) and latent heat flux (λE), integrated across the global land surface – arguably the most fundamental of all carbon-cycle and hydrological quantities – is wide, with many modelled values falling outside of accepted, observationally based ranges (Fig. 5). The problem here is not properly characterized as “uncertainty”. It is rather that many models are certainly incorrect in their representation of the recent past.

It has become recognized across the community of land surface and vegetation modellers that (a) multiple observational constraints are possible, and (b) more systematic application of these constraints is needed to improve confidence in land-surface modelling. Recent reviews (Luo et al., 2012; Foley et al., 2013) and proof-of-concept studies (Randerson et al., 2009; Kelley et al., 2013; Piao et al., 2013) have promoted the concept of model benchmarking against a range of carbon-cycle and hydrological indicators. This is a welcome development. But benchmarking is not a panacea, and there are limits to the extent to which the routine application of observational data sets and data–model comparison metrics can constrain models. Some aspects also need close attention to developments in process understanding, e.g. experimental studies of CO₂ effects on plants (Ainsworth and Long, 2005), or the effects of land-use changes on catchment hydrology (e.g. Siriwardena et al., 2006). Increased confidence in model performance can be achieved through the evaluation of specific assumptions embedded in models against experimental data (Medlyn et al., 2015).

Attention also needs to be paid to model structure, and especially to the way in which natural variability and heterogeneity in biological and physical quantities is represented. It is still common practice in LSMs and DGVMs for highly variable quantities to be represented by a single-valued parameter. For example the hydrological properties of soils are usually assumed either globally constant, or assigned a constant value for each of a small number of soil texture classes; and in any case assigned a constant value across each model grid cell. Biological properties such as leaf photosynthetic capacity have been treated analogously. Many models assign constant biological parameter values within each of a small number of plant functional types (PFTs) even though up to 75% of the observed variation in some important plant traits occurs within PFTs (Kattge et al., 2011). Such devices have the potential to generate artefacts, which should be identifiable as a systematic failure to meet benchmarks. In Sect. 5 we discuss examples of an alternative general approach that appears to yield more robust results.

Figure 4. Simulated land carbon uptake to 2100 under a “high-end” global warming scenario, as projected by global models in the IPCC Third Assessment Report (TAR), Fourth Assessment Report (AR4) and Fifth Assessment Report (AR5). The cross represents the mean of the models included in each assessment.

Figure 5. Mean annual gross primary production (Pg C a⁻¹) and evapotranspiration (mm a⁻¹) from the global land surface during 1901–2010, as simulated by 12 Earth system models in the IPCC Fifth Assessment Report. The grey lines represent upper and lower limits based on observations.
4 Complexity vs. robustness

As more processes continue to be identified and included in LSMs, the almost universal tendency is for LSMs to become more and more complex. A worrying side-effect is the progressive introduction of more model parameters with (commonly) substantially uncertain values. Moreover, complexity can conceal lack of rigour, because it becomes progressively easier to fit observations as more parameters are introduced. Thus, increasing complexity can mask a lack of understanding, resulting in a situation whereby models are tuned to perform well at standard tests but produce widely divergent results when projected beyond the domain of calibration. This seems to be precisely the situation currently observed with coupled carbon-cycle–climate models, as reported in AR5 (Ahlström et al., 2012; Anav et al., 2013; Arora et al., 2013; Jones et al., 2013; Todd-Brown et al., 2013; Ciais et al., 2014). Although it seems reasonable to expect that a model including a larger subset of processes that are known to be important should be more realistic than a simpler model, increases in reliability and robustness by no means automatically follow.

Comparative studies have shown that indeed, complexity in land surface models has not generally improved their reliability (e.g. Desborough and Pitman, 1998). Furthermore, there is no point in achieving sophistication in one set of processes while retaining simple empiricism in another. Complexity needs to be balanced. This is not a precisely defined principle, but it is an important practical one (Smith et al., 2013). We suggest that there is often a trade-off between complexity and robustness, and that robustness is more important than (often spurious) precision. Whereas the representation of a complex system cannot be achieved in a simple model, it seems of paramount importance that complexity is dealt with in a carefully controlled manner that minimizes the scope for over-fitting and thus for the spurious impression of predictive skill.

5 Stochastic parameterization

Stochastic (or statistical) parameterization has gained considerable traction in the atmospheric modelling community, where it has been shown to yield improved robustness and to reduce model artefacts in the numerical representation of weather processes (e.g. Palmer, 2012; Arnold et al., 2013). Stochastic parameterizations represent one or more model parameters as a statistical distribution of values. Atmospheric modelling differs from land-surface modelling in that the equations describing weather processes are inherently chaotic, requiring ensembles of simulations to achieve probabilistic forecasts; implementing a stochastic parameterization in this context can be done by allowing ensemble members to differ in the assignment of parameter values. The equations describing carbon and water cycle processes at and below the land surface are in principle deterministic, in a given environment (Xia et al., 2013). However, the land surface – in contrast with the atmosphere – is heterogeneous at spatial scales down to metres and below, and this heterogeneity cannot be explicitly resolved for the purposes of large-scale modelling. Some form of parameterization is required. Similarly, the ecosystem consists of species with a range of properties, whose aggregate behaviour is not accurately represented by the behaviour of a single species; but a complete enumeration of species and their functional properties would be entirely impractical. As in the atmosphere, the processes represented can be highly non-linear, so that the mean behaviour of the system is not satisfactorily captured by its behaviour at the mean values of the system’s parameters. This is a general property of non-linear systems. Stochastic parameterizations get around this difficulty, and they can often be implemented in a computationally efficient way, avoiding the need for multiple model runs by including calculations on probability density functions within a single realization of the model.

5.1 Hydrological examples

Because runoff is the residual of two relatively large quantities (precipitation vs. evapotranspiration and changes in soil water storage), and because there are no direct observations of evapotranspiration over large areas, streamflow data continue to have a great potential to be used to evaluate LSMs’ simulation of land–atmosphere latent heat and water vapour exchange. (This situation is evolving as improved methods for deriving evapotranspiration from remotely sensed measurements are developed: see Mueller et al., 2013.) Many LSMs fail to generate realistic temporal distributions of streamflow, limiting the potential for such data to be used to test and constrain LSMs. The fundamental problem is that the pointwise generation of runoff is a threshold process (compounded by other highly non-linear properties, including the relationship between hydraulic conductivity and soil water potential) and soil and topographic properties are highly variable. Representing this system by a single “typical” soil profile results in too sharp a transition between high and low flows.

An effective solution to this problem was embedded in the VIC (which stands for “Variable Infiltration Capacity”) LSM (Liang et al., 1994, 1996a) in which the subgrid-scale spatial variabilities of both soil moisture capacity and potential infiltration rate are represented by statistical distributions (Liang and Xie, 2001). The impact of subgrid-scale variability of precipitation is also considered (Liang et al., 1996a). These aspects of variability have significant consequences for the grid-cell total values of the components of the water budget, which are better modelled as a result. VIC has been widely used for land-surface and hydrological impact studies. The soil-moisture capacity curve (a statistical distribution) used for the saturation-excess surface-runoff parameterization in
VIC has been implemented in the ISBA (Habets et al., 1999) and SEWAB (Mengelkamp et al., 1999) LSMs. VIC has been used as a tool to provide retrospective global surface water flux fields (Nijssen et al., 2001). The runoff parameterization of VIC has also been implemented in the Community Land Model (CLM4VIC: Li et al., 2011).

The development of VIC recognized that heterogeneity of land-surface properties is ubiquitous on all spatial scales, down to metres and below. Therefore increasing spatial resolution, tiling, grid nesting and similar devices cannot solve the problem of heterogeneity. Instead, VIC represents subgrid-scale heterogeneity statistically, taking into account the spatial autocorrelation properties as well as variability per se. VIC cannot provide location-specific information on fluxes within each grid cell, but this does not matter, because the objective is only to provide robust information integrated across the grid cell.

Liang and Guo (2003) showed that LSMs such as ISBA and VIC, which explicitly represent the subgrid-scale spatial variability of soil, vegetation, and/or atmospheric forcings, can be less sensitive to the choice of parameter values and thereby produce more robust results, and several other studies have supported this conclusion (e.g. Liang et al., 1996b, 2004; Koren et al., 1999; Li et al., 2011). VIC is insensitive to the assumption of different precipitation distributions within the precipitation-covered area (e.g. Liang et al., 1996b) compared to other LSMs that treat soil properties as invariant (Pitman et al., 1990), and is robust with respect to changes in grid resolution and selection of parameter values (Liang et al., 2004).

A parallel approach has been applied to the routing of streamflow by Wen et al. (2012). This routing scheme, an extension of the one proposed by Guo et al. (2004), applies a statistical distribution for the overland flow path. It is different in several respects from other commonly used routing schemes. Runoff from a grid cell is allowed to exit in multiple directions and a tortuosity coefficient is used to account for geomorphic properties such as channel slope and length. The flow network differentiates explicitly between overland and river flows. The scheme as implemented by Wen et al. (2012) was found to dramatically reduce the dependence of the routing model on the time step, and to produce good results for hourly flows (needed, for example, for flood prediction) where the previous, deterministic parameterization had failed altogether.

DGVMs, even when used for water resources applications, have not generally included parameterizations of land-surface physical variability. However, the inclusion of such a parameterization can greatly improve the hydrological outputs of DGVMs (e.g. Li and Ishidaira, 2011). Exactly why stochastic parameterizations work so well in the context of real landscapes is a research question greatly in need of further study. However, it is worth noting that the statistical properties of landscapes are by no means arbitrary, but are predictable in principle based on the nature of erosion processes (e.g. Turcotte, 2007; Saeki and Okamura, 2010), presumably leading to commonalities that can be exploited for modelling.

5.2 A biological example

Gross primary production (GPP, the space–time integral of carbon uptake by photosynthesis) is the basis of all plant growth. Its global total value is reasonably well constrained by observations (Wang et al., 2014). There is a close coupling between GPP and transpiration, because stomatal opening and closure regulates both CO$_2$ uptake into and water loss out of leaves. Adequate estimation of GPP in the third-generation LSMs is therefore important for modelling the hydrological cycle as well as the carbon cycle. Some of the parameters of photosynthesis (the in vivo enzyme kinetic constants and their temperature responses) can be regarded as constant and well known for global modelling purposes, but others – notably the maximum rate of carboxylation, $V_{\text{cmax}}$, and at least one parameter characterizing the relationship between stomatal conductance and vapour pressure deficit – vary greatly, both within and among species. The usual approach to provide values of these variables in LSMs has been to draw on literature sources to estimate values of each parameter, with the parameters thereby treated as constant (within PFTs) and independent of one another.

There has been little systematic investigation of the consequences of these assumptions. However, just as the representation of hydrological responses can be improved by accounting for the variation and autocorrelation of physical properties within the landscape, it seems likely that the representation of CO$_2$ uptake could be improved by accounting for the variation and covariation of ecophysiological properties within the community of species that carry out photosynthesis.

A vast amount of empirical work during the past decade has gone into the compilation of relevant trait measurements from many plant species (see Wright et al., 2004; Kattge et al., 2011), so the single-value approach can no longer be justified by the paucity of available data (as was the case during the early years of LSM development). In addition to the large variation within PFTs (Kattge et al., 2011), a key finding of this research has been that the parameters, far from being independent, show correlations, so that the variation among species can be collapsed into a few dimensions. One of these dimensions is the so-called leaf economics spectrum, relating photosynthetic rates, leaf longevity and specific leaf area (Wright et al., 2004). Although there has been criticism of the presentation of the leaf economics spectrum, centring on the existence of necessary correlations among various combinations of measurements, its existence and biological significance are not in any doubt (e.g. Lloyd et al., 2013).

In a typical LSM representation, GPP depends on canopy leaf area index and $V_{\text{cmax}}$. Canopy leaf area index is modelled as a function of the fraction of net primary production
allocated to leaves and of the leaf lifespan ($\tau$ in years), and $V_{\text{cmax}}$ is modelled as a function of leaf nitrogen per unit leaf area – i.e. the product of leaf nitrogen concentration ($n$ in g N g$^{-1}$) and leaf mass per area ($m$ in g m$^{-2}$). Field observations from over 50,000 plant species show that leaf lifespan and leaf mass per area are positively correlated, while both are negatively correlated with leaf nitrogen concentration (Wright et al., 2004). Using the CABLE LSM (Kowalczyk et al., 2006; Wang et al., 2010; Wang, 2011), Wang et al. (2012) calculated the global mean and standard deviation of modelled GPP using two groups of 500 randomly sampled sets of the three leaf traits $n$, $\tau$ and $m$ with their observed means and standard deviations. One group also applied the observed covariances of the traits while the other group assumed zero covariance. Simulated global GPP was found to vary from 115 to 170 Gt C a$^{-1}$ when the three model parameters were varied independently. Including covariances did not change the mean GPP, but reduced its standard deviation by 28%, indicating that the observed trait correlations help to constrain the modelled value of global total GPP.

This analysis by Wang et al. (2012) represents a first step towards the realistic inclusion of plant trait variability and correlation patterns in LSMs. The adaptive DGVM approach (Scheiter and Higgins, 2009) represents a somewhat different implementation of stochastic parameterization of plant traits at the continental scale. The general idea that the functional diversity of plants should be represented by continuous trait variation, rather than by a small number of PFTs with fixed characteristics, has been repeatedly mooted (e.g. Kleidon, 2007; van Bodegom et al., 2012). Key to this approach is the idea that functional convergence (the achievement of similar, optimized large-scale fluxes by diverse communities of plants differing in phylogeny) is a consequence of biodiversity, with environmental selection and competition ensuring that niches are filled.

This idea also has the potential to simplify the modelling of GPP and eventually NPP, which is a key quantity for the terrestrial carbon cycle. For example, Wang et al. (2014) have shown that a model explicitly derived from optimality considerations – the least-cost hypothesis of Wright et al. (2003) and Prentice et al. (2014), and the co-limitation or coordination hypothesis (e.g. Maire et al., 2012) – can predict global patterns of forest GPP without the need for PFT-specific parameters. The same has not yet been done for NPP and biomass growth. But the least-cost hypothesis also makes explicit predictions about respiration costs; together with recent findings of general relationships between carbon use efficiency and soil nutrient status (Vicea et al., 2012; Fernández-Martínez, 2014), these predictions are likely to provide the basis for an equally general model of NPP.

6 Towards next-generation models

Figure 6 presents a view of what next-generation LSMs might look like. Key developments needed to make this level of complexity tractable are the implementation of multiple constraints; the use of data assimilation; and the application of stochastic parameterization as discussed above.

6.1 Bounding complexity: the use of multiple constraints

There are encouraging signs that ecologists and ecophysiologists, atmospheric scientists and hydrologists are beginning to work together to improve understanding of large-scale ecosystem and landscape processes, and to identify and quantify the processes that need to be included in LSMs. For example, recognizing the role of deep roots in the function of the soil–plant–atmosphere continuum, researchers are now begin to investigate “new” processes including hydraulic redistribution (e.g. Lee et al., 2005; Amenu and Kumar, 2008; Li et al., 2011; Wang, 2011; Quijano et al., 2012; Luo et al., 2013; Prentice and Cowling, 2013), plant water storage (e.g. Luo et al., 2013), surface water and groundwater interactions (e.g. Winter, 2001; Gutowski et al., 2002; York et al., 2002; Liang et al., 2003; Maxwell and Miller, 2005; Yeh and Eltahir, 2005; Liang et al., 2003; Fan et al., 2007; Niu et al., 2007), and the interactions among these processes and with other existing processes in current LSMs (e.g. Luo et al., 2013). Further new developments include consideration of the relevance of agriculture, wetlands and lakes for the aggregate behaviour of the land surface (e.g. Rosnay et al., 2003; Ringeval et al., 2012; Weble et al., 2012; Drewniak et al., 2013).

With these aspects adding ever-increasing complexity, however, a new modelling strategy is required to ensure that the uncertainties do not spiral out of control as more and more uncertain parameters are introduced. The key lies in ensuring that physical and biological constraints are identified, and explicitly embedded in models. The application of observational constraints (benchmarking against multiple types of observations) routinely during model development is necessary, but not sufficient.

The key principle applied in the recent development of the VIC+ model (Luo et al., 2013) is to enforce multiple constraints on each process, as far as possible, to reduce the number of free (or highly uncertain) parameters in the model. The prototype for this approach was the realization that stomatal conductance to water vapour – which, when combined with leaf area index, is the largest land-surface control on the latent heat flux in vegetated landscapes – must conform (on a fast time scale of seconds) to the same equations (apart from a factor 1.6, relating the molecular diffusivities of water vapour and CO$_2$) that describe how stomatal conductance to CO$_2$ responds to environmental signals. This equality continues to hold even if stomatal conductance is reduced, and/or
Figure 6. Hypothetical schematic of “next-generation” LSMs that will combine the desirable features of previous models, with the addition of surface and subsurface hydrological routing schemes and representations of vegetation dynamics. Model–data fusion and data assimilation will allow effective use of observations from different platforms. Experience suggests that it will be a major challenge to achieve such a complex, realistic representation of land-surface biology and hydrology without loss of reliability and robustness. The application of multiple physical and biological constraints, and the judicious use of stochastic parameterization for subgrid-scale processes, are advocated here as important tools for next-generation model development.

Figure 7. The computational representation of soil–plant–atmosphere water and carbon fluxes in the VIC+ model. Consistency between carbon and water exchanges across the leaf surface, and between water transport from the soil, through plant transport tissues and into the boundary layer, are enforced by means of an iterative algorithm. Plant hydraulic properties (via the Ohm’s law analogy) and stomatal responses thus simultaneously constrain both transpiration and assimilation. Rectangles indicate calculation processes; parallelograms represent variables. From Luo et al. (2013).

photosynthetic capacity inhibited, in response to soil drying (Tuzet et al., 2003; Zhou et al., 2013). Moreover, the rate of photosynthesis implied by the CO$_2$ concentration difference across the stomata must be equal to the rate of photosynthesis implied by the incident photosynthetic photon flux density and key photosynthetic parameters ($V_{c_{\text{max}}}$ and $J_{\text{max}}$). These insights were essential for the inclusion of coupled carbon and water exchanges in the third-generation LSMs (e.g. Collatz et al., 1991). But these are not the only relevant constraints. Allowing for small, but non-zero, water storage, the rate of evaporation at the leaf surface must be equal to the rate of water flow through the xylem; which in turn, following the Ohm’s law analogy for water flows, must be equal to the product of plant hydraulic conductance and the water potential difference between the soil and the leaves. This constraint allows transpiration to be controlled by both the soil water potential of the root zone and the atmospheric conditions simultaneously, mediated by measurable plant characteristics (Tuzet et al., 2003). Figure 7 summarizes how the stomatal and hydraulic constraints are combined in VIC+ to determine the transpiration rate.

VIC+ also represents the influence of soil water potential (via its effect on transpiration, and thus leaf water potential) on stomatal conductance, according to the model of Tuzet et al. (2003) which in turn built on pioneering work by Cowan (1965). The calculation of CO$_2$ assimilation in the model is constrained as a consequence of the interplay of the stomatal and biochemical limitations simultaneously, taking into account the effect of soil moisture signalling, by way of computing the CO$_2$ concentration within the leaf. If transpiration is appropriately represented by $E_{tr1}$ and $E_{tr2}$ (Fig. 7) then these two quantities must converge, as must the two rates $A_{n1}$ and $A_{n2}$ (also shown in Fig. 7) representing CO$_2$ uptake.

The constraints discussed above pertain to physically necessary relationships between fluxes, arising from the architecture of leaves and plants. Potentially, many additional constraints may arise due to natural selection in biological systems, which acts to eliminate “ineffective” combinations of traits, even if they are not directly physically linked. The leaf economics spectrum provides one such set of constraints. The least-cost hypothesis introduced by Wright et al. (2003) and elaborated by Prentice et al. (2014) provides another, potentially powerful constraint, as it leads to an independent specification of the leaf-internal CO$_2$ concentration as calculated in Fig. 7. The co-limitation (or coordination) hypothesis further leads to a prediction of both photosynthetic rate (given leaf temperature and internal CO$_2$ concentration) and $V_{c_{\text{max}}}$ as a function of light availability (Dewar, 1996; Haxeltine and Prentice, 1996; Maire et al., 2012). The resistance to diffusion of CO$_2$ in the mesophyll, between the intercellular spaces and the chloroplasts where photosynthesis is carried out, is another important constraint.
out, is often ignored but can be substantial, and has implications for the strength of CO₂ fertilization (Sun et al., 2014). Again, there is an over-riding physical constraint, i.e. the flux of CO₂ to the chloroplasts must match the net flux of CO₂ into the leaves. \( V_{\text{max}} \) no longer needs to be a PFT-specific parameter but can be predicted dynamically from environmental variations. Moreover the strong relationship between leaf nitrogen and \( V_{\text{cmax}} \) provides a natural way to predict plant nitrogen demand, a key quantity in determining how plants allocate carbon to different functions. With consideration of biologically optimized constraints, we are optimistic that the number of unknown or poorly constrained parameters describing the controls of CO₂ and water exchange by plants can be greatly reduced.

### 6.2 Optimizing model performance: the potential of data assimilation

Obtaining best estimates of parameters, given a set or multiple sets of observations, is one of the goals of data assimilation (e.g. Moradkhani et al., 2005a; Qin et al., 2009; Montzka et al., 2011; Vrugt et al., 2013). Data assimilation has evolved from Newtonian “nudging” to more comprehensive approaches including various flavours of traditional, extended, ensemble Kalman filtering, variational data assimilation using the adjoint method, and the particle filtering method (e.g. Houser et al., 1998; Walker and Houser, 2001; Reichle et al., 2002a; b; Margulis et al., 2002; McLaughlin, 2002; Crow and Wood, 2003; Montaldo and Albertson, 2003; Moradkhani et al., 2005a; b; Pan and Wood, 2006; Qin et al., 2009; Montzka et al., 2011; Vrugt et al., 2013). Parada and Liang (2004) developed a new spatial data assimilation framework, an extension of the Multiscale Kalman Smoother-based (MKS-based) framework (Chou et al., 1994; Fieguth et al., 1995; Luettgen and Willsky, 1995; Kumar, 1999). This framework is innovative in the way it accounts for error propagation, dissimilar spatial resolutions, and the spatial structure within which the distribution of the data is considered. Concepts from this framework have been adopted in several other data assimilation studies (e.g. Parada and Liang, 2008; Pan et al., 2009; Lannoy et al., 2010).

Techniques for data assimilation are thus an active research area. To an even greater extent than for model evaluation and benchmarking, however, the routine use of data assimilation is far from being common practice. It has been stated a number of times that data assimilation should be a standard part of model development. More work is needed to develop generic schemes that would allow data assimilation to be applied to any model, and to set up data sets and protocols for doing so.

Data assimilation, when used to optimize parameter values in a model, is valuable above all because it can potentially reveal whether or not a particular model structure is capable of generating the observed patterns. In normal practice, if a model fails a benchmark test, this does not necessarily indicate that the model is incorrectly specified; it could simply mean that the parameter values in the model are incorrect. If the model fails after assimilation of the relevant data set, however, this may be a strong indication that some structural aspect of the model needs improvement.

Data assimilation confronts a number of practical difficulties. Here we identify three issues that require further research for their satisfactory resolution.

1. High computational demand. Investigators have to choose between gradient-based methods and “brute-force” ensemble simulation (Wang et al., 2009). Ensemble simulations are computationally extremely intensive and can easily become infeasible for global LSMs with several hundred parameters. Gradient-based methods use adjoint codes or finite-difference methods to compute the gradients that are required for optimization (Rayner et al., 2005). The gradient-based approach is many times more efficient than ensembles whenever a large number of parameters are to be optimized. However, adjoint code needs to be generated afresh whenever the model code is modified (Kaminski et al., 2013).

2. Maintaining mass and energy conservation in state assimilation. Compared to offline ecosystem models, one of the advantages of global LSMs is that they enforce the conservation of water, energy and carbon. However many state assimilation techniques do not automatically enforce conservation laws, and need to be modified accordingly. It has yet to be fully explored how such modification affects the parameter estimation process.

3. Quantifying uncertainties in multiple data sets for parameter estimation. Because state-of-the-art LSMs typically include processes with time constants ranging from hours to decades or beyond, multiple data sets with different characteristic temporal and spatial scales are needed to constrain all the model parameters. However the uncertainties of multiple data sets and how those uncertainties vary in space and time are poorly quantified in many cases – introducing an element of subjectivity into the analysis. This problem has been discussed by Raupach et al. (2005) and Wang et al. (2009). A general solution has yet to be found.

### 7 Concluding remarks

Substantial progress has been made in the development of LSMs since Manabe’s pioneering work. The models will continue to evolve. They are already complex. They will become inevitably more complex as they come to represent (a) a more complete description of the set of key processes that determines the exchanges of materials and energy between the atmosphere and the underlying surface and subsurface, for example including surface and groundwater interactions,
sediment transport, and biogeochemical interactions of the carbon, nitrogen and phosphorus cycles; (b) subgrid-scale spatial variability, reflecting the natural diversity of ecosystems and landscapes; and (c) processes requiring high temporal resolution: notably flooding, a key issue in changing climate.

Process understanding continues to evolve, both in biology and in hydrology. At any one time, different models may reasonably differ in the explicit assumptions they make about key processes. This is unavoidable. We suggest that it is also desirable. Global models should incorporate explicit hypotheses about processes, and they are the tool that should allow these hypotheses at the process level to be tested against large-scale observations. Realization of this vision, however, will require teamwork: people with different disciplinary knowledge will need to work together with increased intensity. This is a prerequisite for LSMs to come into their own, as tools for discovery and improved quantitative understanding of the fundamental laws that control energy, water and carbon cycling between the atmosphere and land.

Observational data sets originating in different disciplines, including remote sensing, atmospheric chemistry, ecophysiology and hydrology, will need to be brought to bear routinely to benchmark models and thereby establish their reliability. Robustness will be achieved through the discovery of general regularities that obviate the need to specify large numbers of poorly known or ill-conditioned parameters, such as (non-existent) universal $V_{cmax}$ values for PFTs, and evaluated over time as a community enterprise facilitated by the open publication and sharing of code. Realism will be assessed not as an over-riding requirement to include every known process, but rather by models' ability to give consistent answers to scientific questions, such as the influence of different aspects of climate, environment and land use on global NPP.

The widening field of applications of models to project the consequences of a changing atmospheric and human environments calls for LSMs to be simultaneously reliable, robust and realistic (the three R's of the title) so that they can be used confidently, in new interdisciplinary contexts, to project consequences and potential policy implications of environmental change for agriculture, biodiversity, public health and human security. A new level of reliability is unlikely to be achieved through “business-as-usual” model development. More robust ways to model key processes are within reach, but will require both further scientific development and new code to be written. Several proposals now exist in the literature for possible community-wide benchmark standards, but progress on this front will require community adoption of such standards. A technical facility will be required to help make comprehensive LSM benchmarking and data assimilation a routine process. It will be challenging, but with determination and collaboration, it can be done.

Acknowledgements. This review originated in invited presentations by I. C. Prentice and X. Liang at the inaugural workshop of the Centre for Australian Weather and Climate (a joint activity of CSIRO and the Bureau of Meteorology) in Melbourne in 2007. The ideas were further developed at a workshop on the Hydrological Cycle held at Dartington Hall in 2008 and supported by the Quantifying and Understanding the Earth System (QUEST) programme of the UK Natural Environment Research Council, and at the ILAMB workshop held at the University of California, Irvine in 2011. I. C. Prentice acknowledges funding from the Australian Terrestrial Ecosystem Research Network (TERN) to the ecosystem Modelling And Scaling infrasTructure (eMAST) facility, and from the Australian Research Council (ARC) Discovery project (with Ian Wright) “Next-generation vegetation model based on plant functional types”. X. Liang acknowledges funding from US Department of Energy grant DEFG0208ER64586 and US National Oceanographic and Atmospheric Administration grant NA09OAR4310168. B. E. Medlyn acknowledges funding from the ARC Discovery project (with I. C. Prentice) “Turning water into carbon: a synthesis of plant water-use efficiency from leaf to globe”. Y.-P. Wang acknowledges Long Hui for providing the data for Figs. 4 and 5. I. C. Prentice and X. Liang jointly drafted the text, with contributions from B. E. Medlyn and Y.-P. Wang. X. Liang created the schematic diagrams. This paper is a contribution to the AXA Chair Programme on Biosphere and Climate Impacts, and the Imperial College initiative on Grand Challenges in Ecosystems and the Environment.

Edited by: M. Heimann

References


Arora, V.: Modelling vegetation as a dynamic component in soil–vegetation–atmosphere transfer schemes and hydrological


Murray, S. J., Foster, P. N., and Prentice, I. C.: Evaluation of global continental hydrology as simulated by the Land-surface Pro-


